

# Partnerarbeit „Bash Skript entwickeln“

## 1 Allgemeines

---

Anstelle einer schriftlichen Prüfung wird die Prüfung in Form einer Praxisarbeit durchgeführt. Prüfungen sind nützlich um Umfang und Tiefe des erworbenen Wissens zu testen. Zweckmässiger ist meiner Meinung nach aber den Inhalt eines Kurses zu festigen, was am besten durch selbstständiges Bearbeiten und Anwenden der Inhalte geschieht.

## 2 Auftrag

---

Es ist ein Skript zu entwickeln, welches in einem Menü die verschiedenen Auswertungen der jeweiligen Aufgaben zur Auswahl anzeigt und durch den Benutzer ausgeführt werden kann. Nach abgeschlossener Anzeige des Auswertungsergebnisses (Lösung) soll wieder das Menü angezeigt werden.

## 3 Generelle Anforderungen

---

- Sämtliche Aufgaben (siehe unten) müssen in einer Skriptdatei implementiert sein.
- Die Skript Funktionalität muss getestet sein (Testprotokoll).
- Das Skript muss gut lesbar und Funktionen zur besseren Code Strukturierung verwenden.
- Das Skript ist nach "Clean Code" Richtlinien zu programmieren (Variablen, Formatierung etc.).
- Das Skript muss fehlertolerant aufgebaut sein, sodass ungültige Benutzereingaben behandelt und eine Fehlermeldung ausgegeben wird.
- Die zur Verfügung gestellten Datenquellen (Dateien) dürfen nicht verändert werden.
- Das Skript muss vollständig dokumentiert sein (Doku im Code).
- Die Voraussetzungen (Umgebung, Parameter etc.) zur fehlerfreien Skript Ausführung müssen detailliert dokumentiert sein.
- Das Skript muss beendet werden können.

## 4 Kurzpräsentation

---

Sie stellen Ihre Ergebnisse mittels einer Kurzpräsentation der Klasse vor, präsentieren die implementierten Auswertungen mittels einer Live Demo und schliessen Sie Ihre Präsentation mit einem kurzen Fazit ab (lessons learned).

**Dauer der Kurzpräsentation : ca. 10-15 min**

## 5 Bewertung

### Arbeit

Bewertung	Punkte
Die geforderte Funktionalität ist implementiert	10
Das Skript liefert korrekte Resultate	2
Die korrekte Funktionsweise ist mit Tests verifiziert	2
Der Umfang des Skriptes entspricht den Anforderungen - Die Komplexität der gebotenen Funktionalitäten ist ausreichend	2
Die Implementation ist strukturiert (clean code)	2
Lösung ist praxistauglich	2
<b>Total</b>	<b>20</b>

### Präsentation der Arbeit

Bewertung	Punkte
Systematischer Aufbau der Präsentation / Inhalt / Medienvielfalt	2
Gestaltung und Lesbarkeit der Folien	2
Skript vollständig behandelt / erläutert (Funktionen)	2
Skript ansprechend / grafisch erläutert	2
Fazit	2
<b>Total</b>	<b>10</b>

Die Bewertung setzt sich aus der Arbeit und der Präsentation zusammen und ergibt eine Schlussnote.  
Maximale Punktzahl: 30

**Notenskala:** 
$$\frac{\text{Erreichte Punktzahl} \times 5}{\text{Max. Punktzahl}} + 1 = \text{Note (auf 1/10 Noten gerundet)}$$

## 6 Aufgabe - Universitäten Ranking

Das Ranking der Universitäten ist zu einer gemeinsamen Aufgabe vieler Institutionen geworden, jede von ihnen schlägt eine andere Rangfolge vor, die auf mehreren gewichteten Kategorien basiert. Beispiele von diesen Ranglisten sind: Webometrics Ranking der Universities der Welt, THES - QS World University Rankings und akademisches Ranking der Weltuniversitäten.

**Datenquelle: universities.csv**



universities.csv

**Bemerkung:** Für die formatierte Ausgabe muss das **csvkit** Programm installiert und verwendet werden (<https://csvkit.readthedocs.io>).

### 6.1 Funktion Datenvorschau (Dataset Preview)

Die Grösse einer vollständigen Datenbasis für die Auswertung liegt im Big Data Bereich. Das Laden der ganzen Datei ist somit nicht möglich. Um sich einen Überblick der Datenstruktur verschaffen zu können ist die Funktion Datenvorschau, welche einige Zeilen am Anfang der Datei formatiert auflistet, bereit zu stellen.

**Spalten der Auswertung: Name, Location, State, Tuition and fees, Undergrad Enrollment, Rank**

### 6.2 Funktion Daten Analyse

Suchen und formatierte Ausgabe sämtlicher Spalten aller Universitäten deren Namen ein Suchwort (Teilzeichenkette) enthält. Das jeweilige Suchwort wird vom Benutzer eingegeben (z.B. College).

**Spalten der Auswertung: Name, Location, State, Tuition and fees, Undergrad Enrollment, Rank**

### 6.3 Funktion: Anteil der Colleges

Ermitteln Sie den prozentualen Anteil der Colleges im Verhältnis zu allen aufgeführten Universitäten in der Auswertungsdatei.

**Ausgabe: percent of colleges = 2.15%**

### 6.4 Funktion: Universitäten eines Bundesstaates anzeigen.

Formatierte Ausgabe aller Universitäten eines Bundesstaates. Der gewünschte Bundesstaat kann vom Benutzer festgelegt werden (Eingabe).

**Spalten der Auswertung: Name, Location, State, Tuition and fees, Undergrad Enrollment, Rank**

### 6.5 Funktion: Universitäten eines Bundesstaates.

Ermitteln Sie die Anzahl der Universitäten pro Bundesstaat und geben Sie die Liste aufsteigend sortiert nach Bundesstaat aus.

**Spalten der Auswertung: Count, Bundesstaat**

## 7 Aufgabe - Facebook Data Mining

Unternehmen möchten verstehen, was Menschen über Themen in Bezug auf ihr Geschäft denken, so dass sie ihre Produkte und Marketing Strategie optimieren können. Eine Möglichkeit, dieses Ziel zu erreichen, besteht darin, die FB-Bestände des Unternehmens zu analysieren.

**Datenquelle: facebook.csv**



**Bemerkung:** Für die formatierte Ausgabe muss das **csvkit** Programm installiert und verwendet werden (<https://csvkit.readthedocs.io>).

### 7.1 Funktion Datenvorschau (Dataset Preview)

Die Grösse einer vollständigen Datenbasis für die Auswertung liegt im Big Data Bereich. Das Laden der ganzen Datei ist somit nicht möglich. Um sich einen Überblick der Datenstruktur verschaffen zu können ist die Funktion Datenvorschau, welche einige Zeilen am Anfang der Datei formatiert auflistet, bereit zu stellen.

**Spalten:**

status\_id,status\_message,link\_name,status\_type,status\_link,status\_published,num\_reactions,num\_comments,num\_shares,num\_likes,num\_loves,num\_wows,num\_hahas,num\_sads,num\_angrys

### 7.2 Funktion Daten Analyse

Ermitteln Sie mit einer Auswertung, wie viele Status (*status\_id*) in jeder Statusart (*status\_type*) vorliegen?

**Spalten der Auswertung: Anzahl Status, Statusart**

### 7.3 Funktion beliebtester Statuseintrag

Listen Sie den beliebtesten Statuseintrag (Summe der Spalte 8 bis 15).

Tipp: **awk** für die Summenberechnung verwenden.

**Spalten der Auswertung: Status\_id,Anzahl**

## 8 Aufgabe - Kriminalstatistik

---

Australien ist das sechstgrösste Land der Welt nach Gesamtfläche. Sobald Sie sich entscheiden ein Land zu besuchen, stellt man sich die Frage, welcher Ort oder Stadt die höchste Sicherheit bietet. In jedem Land hat jede Stadt ihre eigenen Eigenschaften, Kriminalitätsrate usw., dasselbe gilt auch für australische Städte.

**Datenquelle: crime.csv**



crime.csv

**Bemerkung:** Für die formatierte Ausgabe muss das **csvkit** Programm installiert und verwendet werden (<https://csvkit.readthedocs.io>).

### 8.1 Funktion: Dateistruktur Analyse

Ermitteln Sie die Anzahl der Zeilen und Spalten in der CSV Datei. Geben Sie das Resultat wie folgt aus:

**Ausgabe: Alle Spaltenbezeichnungen, Anzahl der Zeilen.**

### 8.2 Funktion: Daten Analyse

Ermitteln Sie aus der Datenquelle das häufigste Verbrechen.

**Ausgabe: Verbrechen, Häufigkeit (Anzahl)**

### 8.3 Funktion: Häufigstes Verbrechen pro Stadt

Ermitteln Sie für eine Stadt das häufigste Verbrechen. Die Stadt (z.B. Sydney) wird vom Benutzer eingegeben.

Tipp: **awk** verwenden.

**Ausgabe: Name des Verbrechens, Anzahl Verbrechen**

### 8.4 Funktion: Beste Stadt in Australien

Ermitteln Sie die beste, d.h. die Stadt mit den wenigsten Verbrechen.

Tipp: **awk** verwenden.

**Ausgabe: Name der Stadt, Anzahl Verbrechen**

## 9 Aufgabe – Meteoritendaten analysieren (Python)

---

In dieser Aufgabe müssen die beiden Python Module (Pandas u. Matplotlib) erläutert und Meteoriten Daten grafisch visualisiert werden. Die Datenquelle wurde von NASA bereitgestellt.

**Datenquelle: meteorite.csv**



Meteorite.csv

**Bemerkung:** Verwenden Sie die **Pandas**- (Einlesen von Daten) und **Matplotlib** Module zur Visualisierung der Daten.

### 9.1 Aufgabe: Pandas

Erläutern Sie wozu das **Pandas** Modul entwickelt wurde und zeigen Sie den Einsatz anhand eines praktischen Beispiels.

### 9.2 Aufgabe: Matplotlib

Erläutern Sie wozu das **Matplotlib** Modul entwickelt wurde und zeigen Sie den Einsatz anhand eines praktischen Beispiels.

### 9.3 Aufgabe: Anzahl Meteoriten / Jahr

Ermitteln Sie aus der Datenquelle die Anzahl an einem Jahr (z.B. 1970) runtergefallenen Meteoriten.

### 9.4 Aufgabe: Grafische Visualisierung der Meteoriten / Jahr

Stellen Sie die Anzahl der runtergefallenen Meteoriten pro Jahr über einem Zeithorizont von 1970-2018 grafisch dar.